AD-A120 108    PITTSBURGH UNIV PA INST FOR COMPUTATIONAL MATHEMATICS--ETC    F/G 12/1
ON THE DISCRETIZATION ERROR OF PARAMETIZED NONLINEAR EQUATIONS.(U)
1982    J P FINK, W C RHEINBOLDT    N00014-80-C-0455
UNCLASSIFIED    ICMA-82-40    NL

(15)

AD A120108

# INSTITUTE FOR COMPUTATIONAL MATHEMATICS AND APPLICATIONS

Department of Mathematics and Statistics

University of Pittsburgh





82 08 23 184

15

Technical Report ICMA-82-40

# On the Discretization Error of Parametrized Nonlinear Equations[1)]

by

James P. Fink and Werner C. Rheinboldt

Institute for Computational Mathematics and Applications
Department of Mathematics and Statistics
University of Pittsburgh
Pittsburgh, PA 15261

1) This work was supported in part by the U.S. Air Force Office of Scientific Research under Grant 80-0176, the National Science Foundation under Grant MCS-78-05299 and the Office of Naval Research under Contract N00014-80-C-0455.

## ABSTRACT

Many applications lead to nonlinear, parameter dependent equations $H(y,t) = y_0$, where $H: Y \times T \to Y$, $y_0 \in \mathrm{rge}\, H$, and the state space $Y$ is infinite-dimensional while the parameter space $T$ has finite dimension. The case $\dim T = 1$ is of special interest in connection with continuation methods. For this case, a general theory is developed which provides for the existence of solution paths of a rather general class of such equations and of their finite-dimensional approximations, and which allows for an assessment of the error between these paths. A principal tool in this analysis is the theory of nonlinear Fredholm operators. The results cover a more general class of operators than the mildly nonlinear mappings to which other approaches appear to be restricted.

# On the Discretization Error of Parametrized Nonlinear Equations[1)]

by

James P. Fink[2)] and Werner C. Rheinboldt[2)]

## 1. Introduction

Nonlinear parameter-dependent equations of the form

$$H(y,t) = y_0 \tag{1.1}$$

arise in connection with many equilibrium problems in science and engineering. Here $y$ and $t$ vary in some state space $Y$ and parameter space $T$, respectively, $H$ is a mapping from $Y \times T$ into $Y$, and $y_0$ is a given point in the range of $H$. Under appropriate conditions, the set of all solutions $(y,t)$ of (1.1) forms a manifold in $Y \times T$ and we wish to analyze the properties of this manifold.

In most cases, the state space $Y$ is infinite-dimensional and the parameter space $T$ finite-dimensional, although infinite-dimensional parameter spaces have been considered as well (e.g., see [1]). Thus, for practical computations, there is a need for introducing finite-dimensional approximations of the equation (1.1). This leads to questions about the relation between the properties of the solution manifolds of the original and the discretized equations and the errors introduced by the approximation. Relatively little has been done so far in this area. Without any claim of completeness, we mention here only the articles [2], [5], [7], [12], [15] which, in different settings, address the a priori estimation of the

1) This work was supported in part by the U.S. Air Force Office of Scientific Research under Grant 80-0176, the National Science Foundation under Grant MCS-78-05299 and the Office of Naval Research under Contract N-0014-80-C-0455.

2) Department of Mathematics and Statistics, University of Pittsburgh, Pittsburgh, PA 15261.

approximation errors in the case of a one-dimensional parameter space. In [3], some a posteriori estimates of these errors are considered for certain two-point boundary-value problems.

The restriction to one-dimensional parameter spaces is not as severe as it might appear. In fact, all available computational methods for analyzing solution manifolds of (finite-dimensional) parametrized equations consist principally of some form of continuation process for the trace of paths on the manifold. Such paths are the solution manifolds of some equation involving a one-dimensional parameter.

In this paper, we present a general theory which provides for the existence of solution paths of a rather general class of nonlinear equations with a one-dimensional parameter, as well as of their finite-dimensional approximations, and which allows for an assessment of the error between these paths. A principal tool in this analysis is the theory of nonlinear Fredholm operators. As in [5], [7], the error estimates are derived from a form of the implicit function theorem. But our results cover a more general class of operators than the mildly nonlinear mappings to which the approaches in [5], [7] are restricted.

## 2. Solution Manifolds and Local Parametrizations

In order to avoid repetition, we shall assume throughout this paper that the following information has been given:

(A)

(i) $X, Y$ real Banach spaces;

(ii) $F: E \subset X \to Y$ a mapping of class $C^r(E)$, $r \geq 1$, on some open subset $E \subset X$.

In finite dimensions, an equation involving an m-dimensional parameter corresponds

to the case $\dim X - \dim Y = m$. The analog in infinite dimensions is the requirement that $\ker DF(x)$ has dimension $m$ for the points $x$ under consideration. More specifically, we are interested in regular points of $F$, that is, those $x \in E$ for which $DF(x)$ maps $X$ onto $Y$. Hence, we assume that the _m-regularity set_

$$R_m(F) = \{x \in E: \dim \ker DF(x) = m, \quad DF(x)X = Y\} \tag{2.1}$$

is nonempty for some $m \geq 1$. Then the restriction of $F$ to $R_m(F)$ is a nonlinear Fredholm operator of index $m$ (e.g., see [4]), and this observation is at the center of our analysis.

_Lemma 2.1_: The sets $R_m(F)$, $m \geq 1$, are either empty or open in $X$.

_Proof_: Suppose that $x \in R_m(F)$. Then by the continuity of $DF$ and a basic perturbation property of Fredholm operators (e.g., see [13, p. 115]) there exists a $\delta > 0$ such that $DF(x+\xi)$ is again a Fredholm map of index $m$ for all $\xi \in X$ with $||\xi|| < \delta$. Moreover, we have $\dim \ker DF(x+\xi) \leq \dim \ker DF(x)$, whence $\dim \operatorname{coker} DF(x+\xi) = 0$ and $\dim \ker DF(x+\xi) = m$.

We are interested in examining the solution set of the equation

$$F(x) = y_o, \quad x \in R_m(F), \tag{2.2}$$

for a fixed $y_o \in F(R_m(F))$, that is, the set

$$M_m(y_o) = F^{(-1)}(y_o) \cap R_m(F). \tag{2.3}$$

For any $x_0 \in R_m(F)$, there exist closed subspaces $V \subset X$ such that $X = V \oplus \ker DF(x_0)$. For any such choice of $V$, the restriction $DF(x_0)|V$ is an isomorphism between $V$ and $Y$, and its inverse

$$A_V = (DF(x_0)|V)^{-1} \in L(Y,V) \tag{2.4}$$

is a right inverse of $DF(x_0)$. With this, equation (2.2) may be written in the form

$$F(t + A_V y) = y_0, \quad t \in \ker DF(x_0), \quad y \in Y, \quad t + A_V y \in R_m(F), \tag{2.5}$$

which corresponds to (1.1) with $Y$ as the state space and $T = \ker DF(x_0)$ as the m-dimensional parameter space.

The basic result about the solution set (2.3) may now be phrased as follows:

Theorem 2.2: Suppose that the information (A) is given and that $R_m(F)$ is non-empty for some $m \geq 1$. Then, for any $y_0 \in F(R_m(F))$, the regular solution set $M_m(y_0)$ is a nonempty, relatively open, m-dimensional $C^r$-manifold in $X$.

Proof: The proof proceeds along the lines of the finite-dimensional analog given in [10]. Let $x_0 \in M_m(y_0)$ and $V$ be a subspace of $X$ such that $X = V \oplus \ker DF(x_0)$. With the corresponding projection $P: X \to \ker DF(x_0)$, we define the mapping $G: E \to Y \times \ker DF(x_0)$ by $G(x) = (F(x), Px)$ for $x \in E$. Then $DG(x)\xi = (DF(x)\xi, P\xi)$, $\xi \in X$, and $DG(x_0)$ is nonsingular. Hence, the inverse function theorem applies and $G$ maps some neighborhood $U \subset X$ of $x_0$ $C^r$-diffeomorphically onto a neighborhood $S$ of $(y_0, Px_0)$. Evidently now, $G$ maps $M_m(y_0) \cap U$ $C^r$-diffeomorphically onto $(\{y_0\} \times \ker DF(x_0)) \cap S$ and, since

$\{y_0\} \times \ker DF(x_0)$ is $C^r$-diffeomorphic to $\mathbb{R}^m$, the result follows.

As mentioned in the introduction, the computational procedures for analyzing the solution manifold $M_m(y_0)$ of (2.2) consist principally of methods for approximating paths on this manifold. Such a path is defined as the solution manifold of some reduced equation defined by a Fredholm operator of index 1 on its regularity set. Therefore, from now on we restrict ourselves to the case when (A) is given and $R_1(F)$ is nonempty. For ease of notation, we will write $R(F) = R_1(F)$ and $M(y_0) = M_1(y_0)$.

We consider first the question of the choice of suitable local parametrizations of the one-dimensional manifold $M(y_0)$. It turns out that a possible choice corresponds to a typical approach used in many continuation procedures.

In line with the earlier discussion leading to (2.5), a <u>local parametrization</u> of $M(y_0)$ at a given point $x_0 \in M(y_0)$ is defined as a triple $\{V,A,z_0\}$ consisting of a closed subspace $V$ of $X$, a linear operator $A \in L(Y,V)$, and a point $z_0 \in X$ such that

$$\begin{aligned} &\text{(i)} \quad V \cap \ker DF(x_0) = \{0\};\\ &\text{(ii)} \quad A \text{ is an isomorphism of } Y \text{ onto } V;\\ &\text{(iii)} \quad z_0 \notin V;\\ &\text{(iv)} \quad X = V \oplus T_0, \quad T_0 = \operatorname{span}\{z_0\}. \end{aligned} \tag{2.6}$$

As indicated in Figure 1, we consider the family of parallel linear manifolds

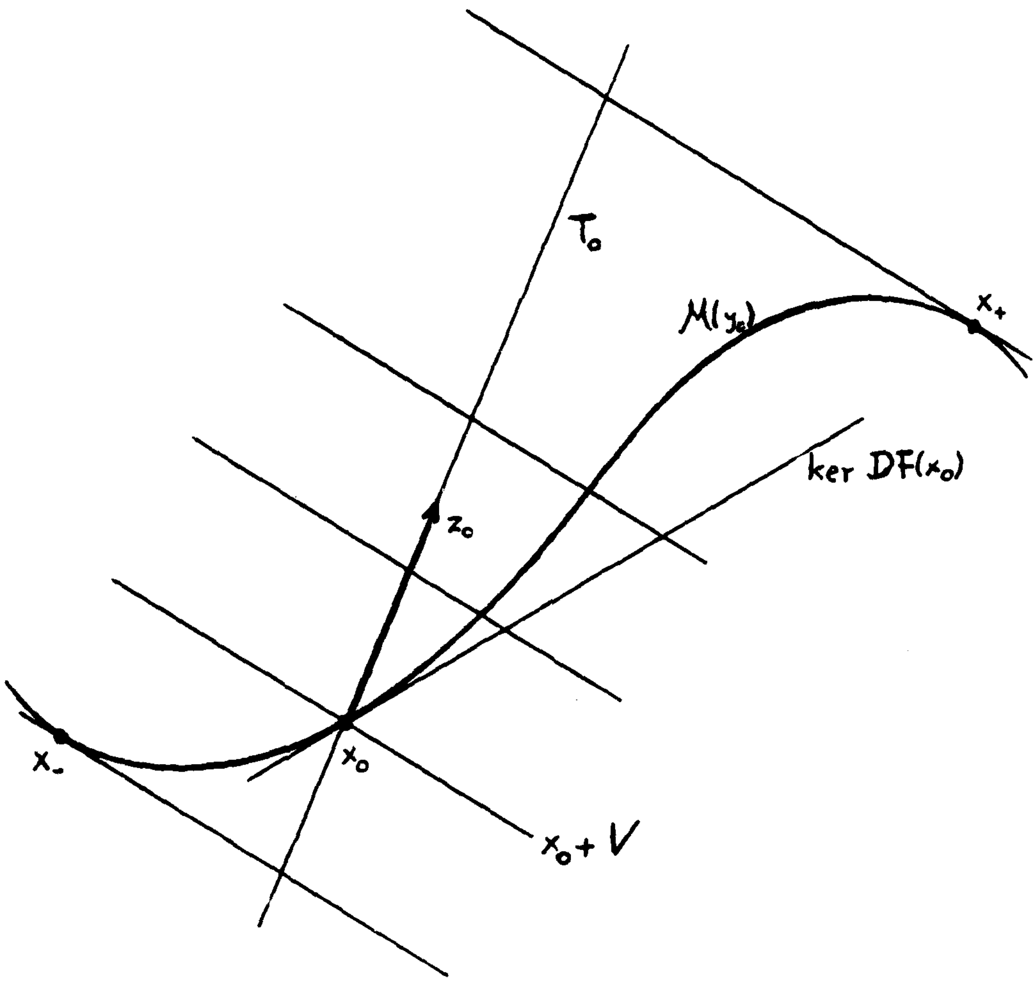


Figure 1

$$x_0 + tz_0 + V, \quad t \in \mathbb{R}^1. \tag{2.7}$$

The conditions (2.6) ensure that these manifolds are transverse to $\ker DF(x_0)$ and hence that, locally near $t = 0$, the solution manifold may be parametrized in terms of $t$. This is the content of the following result:

Theorem 2.3: Let (A) be given, $R(F) \neq \phi$, and $\{V,A,z_0\}$ a local parametrization of $M(y_0)$ at a point $x_0 \in M(y_0)$. Then there exists an open interval $J \subset \mathbb{R}^1$, $0 \in J$, an open neighborhood $U \subset X$ of $x_0$, and a unique $C^r$-map $\eta: J \to Y$, such that

$$M(y_0) \cap U = \{x \in X: x = x_0 + tz_0 + A\eta(t), \quad t \in J\}. \tag{2.8}$$

Proof: Let $B: \mathbb{R}^1 \times Y \to X$ denote the affine mapping $B(t,y) = x_0 + tz_0 + Ay$, $t \in \mathbb{R}^1$, $y \in Y$, and define $H: B^{(-1)}(E) \to Y$ by

$$H(t,y) = F(B(t,y)) - y_0, \quad (t,y) \in B^{(-1)}(E). \tag{2.9}$$

Then $H$ is of class $C^r$ on $B^{(-1)}(E)$ and $H(0,0) = 0$. Moreover, since $A$ is an isomorphism of $Y$ onto $V$ and $DF(x_0)$ maps $V$ one-to-one onto $Y$, it follows that $D_yH(0,0) = DF(x_0)A$ is an isomorphism of $Y$. Therefore, by the implicit function theorem, there exists an open interval $J \subset \mathbb{R}^1$, $0 \in J$, a neighborhood $S \subset Y$ of $y = 0$, and a unique $C^r$-function $\eta: J \to S$ such that

$$H(t,\eta(t)) = F(x_0 + tz_0 + A\eta(t)) - y_0 = 0, \quad t \in J.$$

With $U = B(J \times S)$, the uniqueness of $\eta$ ensures that (2.8) holds.

In the standard continuation methods, $z_0$ defines the predictor line and the corrector produces the step $A\eta(t)$ from the predicted point $x_0 + tz_0$ to the point $x_0 + tz_0 + A\eta(t)$ on $M(y_0)$. A particular local parametrization is obtained if we choose any nonzero vector $z_0 \in \ker DF(x_0)$ and the map $A_V$ of (2.4) for any closed subspace $V \subset X$ such that $X = V \oplus \ker DF(x_0)$. This choice corresponds essentially to the pseudo-arclength parametrizations proposed by various authors (e.g., see [8]).

The question arises as to how far a particular local parametrization $\{V,A,z_0\}$ can be extended. For this, note that the set

$$\dot{A} = \{x \in E\colon DF(x)A \text{ is an isomorphism of } Y\} \tag{2.10}$$

is certainly nonempty and open. With this observation, a generalization of a result in [9] may be phrased as follows:

<u>Theorem 2.4</u>: Under the conditions of Theorem 2.3, let $M_\infty \subset X$ denote the maximal connected subset of $M(y_0) \cap \dot{A}$ which contains $x_0$. Then there exists an open interval $J_\infty \subset \mathbb{R}^1$, $0 \in J_\infty$, and a $C^r$-function $\eta_\infty\colon J_\infty \to Y$ such that

$$M_\infty = \{x \in X\colon\ x = x_0 + tz_0 + A\eta_\infty(t),\ \ t \in J_\infty\}.$$

<u>Proof</u>: Since $X = V \oplus T_0$, each $x \in X$ can be written uniquely in the form $x = x_0 + tz_0 + Ay$ for $(t,y) \in \mathbb{R}^1 \times Y$. Thus, each $x_1 \in M_\infty$ may be expressed uniquely as $x_1 = x_0 + t_1z_0 + Ay_1$ for $(t_1,y_1) \in B^{(-1)}(E)$, where $B$ is defined in the proof of Theorem 2.3. Let $H\colon B^{(-1)}(E) \to Y$ be given by (2.9). Then, for any $x_1 \in M_\infty$, the derivative $D_yH(t_1,y_1) = DF(x_1)A$ is an isomorphism of $Y$, and we may apply the implicit function theorem to ensure the existence of an open interval $J_{x_1} \subset \mathbb{R}^1$, $t_1 \in J_{x_1}$, a neighborhood $S_{x_1} \subset Y$ of $y_1$, and a unique

$C^r$-function $\eta_{x_1}: J_{x_1} \to S_{x_1}$ such that $H(t,\eta_{x_1}(t)) = 0$ for $t \in J_{x_1}$. Consider the projection $\pi: \mathbb{R}^1 \times Y \to \mathbb{R}^1$ and its restriction to $N_\infty = B^{(-1)}(M_\infty)$. Then, for each $x_1 \in M_\infty$, and hence each $(t_1,y_1) \in N_\infty$, there exists an open interval $J_{x_1} \subset \mathbb{R}^1$ containing $t_1$ such that $\pi_\infty = \pi|N_\infty$ is a one-to-one mapping of $\pi_\infty^{(-1)}(J_{x_1})$ onto $J_{x_1}$. In fact, we have

$$\pi_\infty^{-1}(t) = (t,\eta_{x_1}(t)),\ t \in J_{x_1}.$$

Thus, $\pi_\infty$ is a local $C^r$-diffeomorphism of $N_\infty$ to $\mathbb{R}^1$. A classical result of topology states that, as a local homeomorphism from the connected Hausdorff space $N_\infty$ to $\mathbb{R}^1$, $\pi_\infty$ must be a homeomorphism from $N_\infty$ onto $J_\infty = \pi_\infty(N_\infty)$. Now with the other projection $\phi: \mathbb{R}^1 \times Y \to Y$ we may define $\eta_\infty: J_\infty \to Y$ by $\eta_\infty = \phi \circ \pi_\infty^{-1}$ and the result follows.

The condition that, for $x \in M_0$, the mapping $DF(x)A$ is an isomorphism of $Y$ means geometrically that a local parametrization $\{V,A,z_0\}$ is valid for the segment of the solution path $M(y_0)$ between the points $x_-$ and $x_+$ closest to $x_0$ at which $x_- + V$ and $x_+ + V$ are tangent to $M(y_0)$ (see Figure 1). In standard terminology, these are the closest limit points of the path with respect to the decomposition $X = V \oplus \text{span}(z_0)$.

## 3. Finite-Dimensional Approximations

In this section, we turn to the formulation of suitable approximate problems for (2.2). In many applications, one has $X = Y \times \mathbb{R}^1$; that is, a particular component of $X$ is identified as a parameter and only the complementary component $Y$ needs to be discretized. We generalize this situation by assuming that a splitting

$$X = Z \oplus T, \quad \dim T = 1 \tag{3.1}$$

has been given together with an operator which relates $Z$ to $Y$: namely,

$$\begin{gathered} Q \in L(X,Y), \quad T = \ker Q, \\ Q|Z \in L(Z,Y) \text{ is an isomorphism.} \end{gathered} \tag{3.2}$$

As in the previous section, suppose that our basic data (A) are given and that $R(F) = R_1(F)$ is nonempty. A discretization of the problem is specified by a set $\{P_h\}$ of projections $P_h \in L(Y)$ of finite rank, indexed by positive $h > 0$, such that

$$\lim_{h \to 0} P_h y = y, \quad y \in Y. \tag{3.3}$$

With the families of subspaces

$$Y_h = P_h Y, \; Z_h = (Q|Z)^{-1} Y_h, \; X_h = Z_h \oplus T, \tag{3.4}$$

the approximate problem is then given by

$$F_h(x) = y_{oh}, \quad x \in R(F_h), \quad y_{oh} = P_h y_o, \tag{3.5}$$

where

$$F_h : E_h \subset X_h \to Y_h, \quad F_h(x) = P_h F(x), \quad x \in E_h = E \cap X_h. \tag{3.6}$$

For ease of discussion, we call the information given by the projections $\{P_h\}$, the subspaces (3.4), and the approximate equations (3.5) and operators (3.6) a <u>basic approximation</u> of the problem (2.2).

The general question we address in this and the next section is whether the solution manifolds of the discretized equations (3.5) approximate the solution manifold $M(y_o)$ of the original equation (2.2) when $h$ tends to zero. More specifically, in this section we consider the existence of solutions of (3.5) and take up the matter of error estimates in the subsequent section.

For the analysis of the approximate problems, it is convenient to extend the discrete operators (3.6) to all of $E \subset X$ as follows:

$$\hat{F}_h: E \to Y, \quad \hat{F}_h(x) = (I-P_h)Qx + P_h(F(x)-y_o), \quad x \in E, \tag{3.7}$$

where $I$ denotes the identity on $Y$. Clearly, $\hat{F}_h$ is a $C^r$-mapping and the following properties hold:

<u>Proposition 3.1</u>:

(i) $\hat{F}_h(x) = 0$ for $x \in E$ if and only if $x \in E_h$ and $F_h(x) = y_{oh}$;

(ii) $D\hat{F}_h(x)X_h \subset Y_h$, $x \in E$;

(iii) $\ker D\hat{F}_h(x) \subset X_h$, $x \in E$;

(iv) $D\hat{F}_h(x) \in L(X,Y)$ is a Fredholm operator of index 1 for all $x \in E$;

(v) $P_h DF(x)X_h = Y_h$ for some $x \in E$ implies that $x \in R(F_h)$.

<u>Proof</u>: The proofs of (i), (ii), and (iii) are straightforward. For example, if $\hat{F}_h(x) = 0$ for some $x \in E$, then $Qx = P_hQx - P_h(F(x)-y_o) \in Y_h$ implies that $x \in X_h$ and hence that $F_h(x) = P_hF(x) = \hat{F}_h(x) - (I-P_h)Qx + P_hy_o = y_{oh}$. The converse of (i), as well as (ii) and (iii), follow analogously. The property

(iv) is a consequence of the fact that $D\hat{F}_h(x) = Q - P_hQ + P_hDF(x)$ is a compact perturbation of the Fredholm operator $Q \varepsilon L(X,Y)$ of index 1. Since $D\hat{F}_h(x)|X_h = P_hDF(x)X_h$, we obtain (v) from (iii) and (iv) if only it can be shown that $\dim\ker D\hat{F}_h(x) = 1$. But, since by assumption $\mathrm{rge}\, P_hDF(x)|X_h = Y_h$ and $\dim X_h = \dim Y_h + 1$, this follows immediately from $\dim\ker P_hDF(x)|X_h + \dim\mathrm{rge}\, P_hDF(x)|X_h = \dim X_h$.

Clearly, any comparison of the solution manifolds of (2.2) and (3.5) must be done locally. Let $x_0 \varepsilon M(y_0)$ be given and suppose that $\{V,A,z_0\}$ is a local parametrization of $M(y_0)$ at $x_0$. This parametrization must relate in a suitable way to the basic discretization introduced above, and the compatibility condition may be expressed in a variety of forms. Rather than go into a detailed discussion of various possible equivalent definitions, we introduce here simply the following technical condition which is relatively easy to verify in many practical situations:

$$\text{(C)} \qquad ||D\hat{F}_h(x_0)Ay|| \geq \gamma||y||, \quad y \varepsilon Y, \text{ for all sufficiently small } h > 0.$$

Here $\gamma > 0$ is a constant independent of $y$ and $h$.

Compatibility conditions of this form typically arise in the consideration of projection methods. For example, condition (C) is related to the notion of stable convergence of linear operators defined and discussed in [14].

The condition (C) can be enforced in a number of ways. For example, by a judicious, although somewhat restrictive, choice of the discretization and parametrization, $D\hat{F}_h(x_0)A$ turns out to be the identity on $Y$ and (C) holds with $\gamma = 1$ for all $h > 0$. As discussed in Section 7, this case includes the so-called reduced-basis technique (e.g., see [11]). On the other hand, as noted at the beginning of this section, in many applications the formulation of the

problem includes a natural splitting of the space $X$, and then (C) becomes in essence a condition on $F$ if we are not prepared to restrict the discretizations and parametrizations. In particular, we show in Sections 5 and 6 several classes of operators $F$ for which (C) is easily verified.

As a first consequence of (C), we obtain the following result:

Proposition 3.2: If (C) holds, then $P_h DF(x_o)X_h = Y_h$ and $x_o \in R(\hat{F}_h)$ for all sufficiently small $h > 0$.

Proof: Evidently, (C) implies that $D\hat{F}_h(x_o)A \in L(Y)$ is an isomorphism. Hence, for $y_1 \in Y_h$, there exists a $y_2 \in Y$ such that $D\hat{F}_h(x_o)Ay_2 = y_1$, and therefore $QAy_2 = P_hQAy_2 - P_hDF(x_o)Ay_2 + y_1 \in Y_h$. From this it follows that $Ay_2 \in X_h$, $(I-P_h)QAy_2 = 0$, and $P_hDF(x_o)Ay_2 = y_1$. The second assertion is now a direct consequence of Proposition 3.1 (v).

In order to prove the existence of solutions of the approximate problems (3.5), we make use of a generalized form of the inverse function theorem. As noted in [5], the usual proof of this theorem provides a Lipschitz condition for the inverse function. In our terminology, the result may be formulated as follows:

Theorem 3.3: Let (A) be given, where, at some $x_o \in E$, (i) $DF(x_o) \in L(X,Y)$ is an isomorphism onto $Y$ with $||DF(x_o)^{-1}|| \leq 1/\gamma$, and (ii) there exists a $\delta > 0$ such that

$$||DF(x)-DF(x_o)|| \leq \frac{\gamma}{2}, \qquad x \in B(x_o,\delta) = \{x \in X: ||x-x_o|| < \delta\} \subset E.$$

Then there exists a unique $C^r$-map $G: B(F(x_o),\gamma\delta/2) \to B(x_o,\delta)$ for which

$F(G(y)) = y$ for all $y \in B(F(x_0),\gamma\delta/2)$ and

$$||G(y_1)-G(y_2)|| \leq \frac{2}{\gamma} ||y_1-y_2||, \quad y_1,y_2 \in B(F(x_0),\gamma\delta/2).$$

In order to prepare the way for the application of this theorem, let $B: Y \to X$ be the affine mapping given by $By = x_0 + Ay$, $y \in Y$, and define the operators

$$H_h: B^{(-1)}(R(F)) \to Y, \quad H_h(y) = \hat{F}_h(By), \quad y \in B^{(-1)}(R(F)). \tag{3.8}$$

Clearly, $H_h$ is of class $C^r$ and $DH_h(y) = (I-P_h)QA + P_h DF(x_0+Ay)A$.

We wish to apply Theorem 3.3 to $H_h$. The two conditions of the theorem are verified in the following lemma:

Lemma 3.4: Let (C) hold. Then $DH_h(0) \in L(Y)$ is an isomorphism of $Y$ with $||DH_h(0)^{-1}|| \leq 1/\gamma$ for all sufficiently small $h > 0$. Moreover, there exists a $\delta > 0$, independent of $h$, such that

$$||DH_h(y)-DH_h(0)|| \leq \frac{\gamma}{2} \quad \text{whenever} \quad ||y|| < \delta. \tag{3.9}$$

Proof: The first part is an immediate consequence of (C). In order to prove (3.9), note first that, from (3.3) and the uniform boundedness principle, it follows that

$$||P_h|| \leq \alpha \quad \text{for all sufficiently small} \quad h > 0 \tag{3.10}$$

with a constant $\alpha > 0$ which does not depend on $h$. Hence, we find that

$$||DH_h(y)-DH_h(0)|| = ||P_h(DF(x_o+Ay) - DF(x_o))A||$$

$$\leq \alpha||A|| \; ||DF(x_o+Ay) - DF(x_o)||,$$

and (3.9) is a direct consequence of the continuity of DF.

With this, we may now apply Theorem 3.3 to $H_h$ for sufficiently small $h > 0$. This ensures the existence of a unique $C^r$-function

$$G_h: \mathcal{B}_o = \mathcal{B}((I-P_h)Qx_o, \gamma\delta/2) \to \mathcal{B}(0,\delta) \tag{3.11}$$

for which

$$\begin{aligned} &\text{(i)} \quad H_h(G_h(y)) = y, \quad y \in \mathcal{B}_o, \\ &\text{(ii)} \quad ||G_h(y_1)-G_h(y_2)|| \leq \frac{2}{\gamma}||y_1-y_2||, \quad y_1,y_2 \in \mathcal{B}_o. \end{aligned} \tag{3.12}$$

Since $\gamma$ and $\delta$ are independent of $h$ and $(I-P_h)Qx_o \to 0$ as $h \to 0$, we see that $0 \in \mathcal{B}_o$ for all sufficiently small $h > 0$. Hence, $y_h = G_h(0)$ satisfies $H_h(y_h) = 0$, and from (3.12) (ii) with $y_1 = 0$ and $y_2 = (I-P_h)Qx_o$ it follows that $||y_h|| \leq (2/\gamma)\; ||(I-P_h)Qx_o||$. Hence, for $x_{oh} = x_o + Ay_h$, we obtain $\hat{F}_h(x_{oh}) = H_h(y_h) = 0$ as well as $x_{oh} \to x_o$ when $h \to 0$. Moreover, Proposition (3.1) (i) implies that $x_{oh} \in E_h$ and $F_h(x_{oh}) = y_{oh}$. In other words, $x_{oh}$ is the desired solution of the approximate problem (3.5) corresponding to $x_o$. These solutions have several interesting properties:

<u>Proposition 3.5</u>: If (C) holds, then $x_{oh} \in R(\hat{F}_h)$ for all sufficiently small $h$.

<u>Proof</u>: Since $R(F)$ is open, we certainly have $x_{oh} \in R(F)$ for all sufficiently small $h$. Moreover, by Proposition 3.2, we also know that $x_o \in R(\hat{F}_h)$. But, in

order to conclude that the $x_{oh}$ ultimately belong to $R(\hat{F}_h)$, we need to use the uniformity implied by (C). For this, we write

$$D\hat{F}_h(x_{oh})A = D\hat{F}_h(x_o)A + (D\hat{F}_h(x_{oh})-D\hat{F}_h(x_o))A$$

and note that, for all sufficiently small $h > 0$, it follows from Lemma 3.4 that $D\hat{F}_h(x_o)A \in L(Y)$ is an isomorphism with $||(D\hat{F}_h(x_o)A)^{-1}|| \leq 1/\gamma$ and, moreover, that, by (3.9) with $y = y_h$,

$$||(D\hat{F}_h(x_{oh})-D\hat{F}_h(x_o))A|| \leq \frac{\gamma}{2}.$$

Therefore, $D\hat{F}_h(x_{oh})A \in L(Y)$ is also an isomorphism of $Y$ and thus $D\hat{F}_h(x_{oh})$ must map $X$ onto $Y$. Since, by Proposition (3.1) (iv), $D\hat{F}_h(x_{oh})$ is a Fredholm operator of index 1, the result follows.

<u>Proposition 3.6</u>: If (C) holds, then $V \cap \ker D\hat{F}_h(x_{oh}) = \{0\}$.

<u>Proof</u>: In the proof of Proposition 3.5, we saw that $D\hat{F}_h(x_{oh})A \in L(Y)$ is an isomorphism of $Y$ for sufficiently small $h$. Hence, if $D\hat{F}_h(x_{oh})v = 0$ for some $v \in V$, then $v = Ay$ for some $y \in Y$, and therefore $D\hat{F}_h(x_{oh})Ay = 0$, whence $y = 0$ and thus also $v = 0$.

We collect our results up to this point in the following theorem:

<u>Theorem 3.7</u>: Let (A) be given and $R(F) \neq \phi$. Suppose that a basic approximation has been chosen and that $\{V,A,z_o\}$ is a local parametrization of $M(y_o)$ at $x_o \in M(y_o)$, where the compatibility condition (C) is satisfied. Then, for all sufficiently small $h > 0$, the approximate problems (3.5) have solutions

$x_{oh} \in R(F_h)$ for which $\lim_{h\to 0} x_{oh} = x_o$. Moreover, for any local parametrization of the form $\{V_h, (P_h DF(x_{oh})|V_h)^{-1}, z_{oh}\}$ of the solution manifold $F_h^{(-1)}(y_{oh}) \cap R(F_h)$ of (3.5) at $x_{oh}$, Theorems 2.2, 2.3 and 2.4 apply.

## 4. Error Estimates

Theorem 3.7 ensures the existence of a solution segment for the local approximate problem (3.5) under the same general conditions on the operator $F$ needed to establish the existence of a solution segment for the full problem (2.2). If error estimates are desired, however, then additional smoothness conditions are required for $F$. Moreover, the parametrizations of the two curves cannot be chosen independently, but must be such as to allow for a specific association of comparable points.

For the derivation of the error estimates, we use a form of the implicit function theorem formulated in [5] which follows from Theorem 3.3.

Theorem 4.1: Let $X,Y,Z$ be Banach spaces, and $\Phi: S \subset X \to Y$ a Lipschitz-continuous mapping with Lipschitz constant $\gamma_2 > 0$ on the subset $S$ of $X$. Let $F: E \subset X \times Y \to Z$ be a $C^r$-mapping on a neighborhood $E$ of $S \times \Phi(S)$ such that (i) for each $x \in S$, $D_yF(x,\Phi(x))$ is an isomorphism of $Y$ onto $Z$ for which $\sup\{||D_yF(x,\Phi(x))^{-1}||: x \in S\} \leq \gamma_0$ and $\sup\{||D_xF(x,\Phi(x))||: x \in S\} \leq \gamma_1$, and (ii) there exists a monotonically increasing function $\ell: \mathbb{R}^1_+ \to \mathbb{R}^1_+$ for which

$$||DF(x,y)-DF(x_0,\Phi(x_0))|| \leq \ell(\rho)\ ||(x,y)-(x_0,\Phi(x_0))||, \quad x_0 \in S,$$
$$(x,y) \in B((x_0,\Phi(x_0)),\rho).$$

Then there are constants $a > 0$ and $b > 0$, depending only on $\gamma_0, \gamma_1, \gamma_2$, and $\ell$, such that, for $\sup\{||F(x_0, \Phi(x_0)||: x_0 \in S\} \le \varepsilon = a/2$, there exists a $C^r$-function $G: \bigcup_{x_0 \in S} B(x_0, a) \to Y$ which satisfies

$$F(x, G(x)) = 0, \quad x \in \bigcup_{x_0 \in S} B(x_0, a),$$
$$G(B(x_0, a)) \subset B(\Phi(x_0), b), \quad x_0 \in S, \tag{4.1}$$

and

$$||G(x) - \Phi(x_0)|| \le C_0(||x - x_0|| + ||F(x_0, \Phi(x_0))||), \quad x_0 \in S, \quad x \in B(x_0, a), \tag{4.2}$$

where the constant $C_0 > 0$ depends only on $\gamma_0$ and $\gamma_1$.

In order to set up a situation in which we can use this theorem, we continue to assume the conditions of Theorem 3.7. By Theorem 2.3, the original problem (2.2) has a solution segment $x(t) = x_0 + tz_0 + A\eta(t)$ defined for $t$ in an open interval $J \subset \mathbb{R}^1$ containing $t = 0$. By the continuity of $DF$, there exists a nontrivial compact subinterval $J_0 \subset J$, $0 \in J_0$, such that

$$||DF(x_0 + tz_0 + A\eta(t)) - DF(x_0)|| \le \beta < \frac{\gamma}{\alpha ||A||}, \quad t \in J_0, \tag{4.3}$$

where $\alpha$ is defined by (3.10) and $\gamma$ by condition (C).

Proposition 3.6 provides a clue how to proceed because it suggests that the parametrization $\{V, A, z_0\}$ of the original curve may also be suitable for the approximate curve. With this in mind, we define again an affine mapping $B: \mathbb{R}^1 \times Y \to X$ by $B(t, y) = x_0 + tz_0 + Ay$, $t \in \mathbb{R}^1$, $y \in Y$, and with it the operator

$$H_h\colon B^{(-1)}(E) \subset \mathbb{R}^1 \times Y \to Y, \quad H_h(t,y) = \hat{F}_h(B(t,y)), \quad (t,y) \in B^{(-1)}(E). \tag{4.4}$$

Clearly, $H_h$ is of class $C^r$ with derivatives

$$D_tH_h(t,y) = D\hat{F}_h(x_o+tz_o+Ay)z_o = (I-P_h)Qz_o + P_hDF(x_o+tz_o+Ay)z_o,$$

$$D_yH_h(t,y) = D\hat{F}_h(x_o+tz_o+Ay)A = (I-P_h)QA + P_hDF(x_o+tz_o+Ay)A.$$

The following three lemmas show that $H_h$ satisfies the conditions of Theorem 4.1.

Lemma 4.2: For sufficiently small $h$ and for each $t \in J_o$, $D_yH_h(t,\eta(t))$ is an isomorphism of $Y$ and we have $\sup\{||D_yH_h(t,\eta(t))^{-1}||: t \in J_o\} \le \gamma_o$, where $\gamma_o$ is independent of $h$.

Proof: The proof proceeds along the lines of Proposition 3.5. In brief, by condition (C), $D_yH_h(0,0) = D\hat{F}_h(x_o)A$ is an isomorphism of $Y$ with $||D_yH_h(0,0)^{-1}|| \le 1/\gamma$. From this, it follows that $D_yH_h(t,\eta(t))$ is an isomorphism for $t \in J_o$ and now (4.3) can be used to complete the proof.

Lemma 4.3: For sufficiently small $h$, we have $\sup\{||D_tH_h(t,\eta(t))||: t \in J_o\} \le \gamma_1$, where $\gamma_1$ is independent of $h$.

Proof: This estimate follows directly from (3.10) and the continuity of $DF(x_o + tz_o + A\eta(t))$ for $t$ in the compact interval $J_o$.

Lemma 4.4: $\lim_{h\to 0} \sup_{t\varepsilon J_o} ||H_h(t,\eta(t))|| = 0.$

Proof: We have

$$\sup_{t\varepsilon J_o} ||H_h(t,\eta(t))|| \leq ||(I-P_h)Qx_o|| + M||(I-P_h)Qz_o|| + \sup_{t\varepsilon J_o} ||(I-P_h)QA\eta(t)||,$$

where $M = \max_{t\varepsilon J_o} |t|$. By (3.3), it follows that $(I-P_h)Qx_o \to 0$, $(I-P_h)Qz_o \to 0$, and $(I-P_h)QA\eta(t) \to 0$ pointwise for $t \varepsilon J_o$. From (3.10) and the continuity of $\eta'(t)$, we obtain, moreover, that $(I-P_h)QA\eta(t) \to 0$ uniformly on $J_o$, which implies the result.

If we assume now that DF satisfies a Lipschitz condition, then we are in a position to apply Theorem 4.1. More specifically, suppose that DF is Lipschitz-continuous on bounded subsets, that is, that

$$||DF(x_1)-DF(x_2)|| \leq \ell(\rho)\ ||x_1-x_2||, \quad x_1,x_2 \varepsilon E, \quad x_2 \varepsilon B(x_1,\rho), \qquad (4.5)$$

with some monotonically increasing function $\ell: \mathbb{R}^1_+ \to \mathbb{R}^1_+$. Then we obtain the following theorem:

Theorem 4.5: Suppose that the conditions of Theorem 3.7 hold and that DF satisfies (4.5). By Theorem 2.3, the original problem (2.2) has a solution segment $x: J \subset \mathbb{R}^1 \to E$ defined on an open interval $J \subset \mathbb{R}^1$, $0 \varepsilon J$. Then there exists a compact subinterval $J_o \subset J$, $0 \varepsilon J_o$, such that, for sufficiently small $h$, the local approximate problem (3.5) has a solution segment $x_h: J_o \to E_h$ and

$$||x(t)-x_h(t)|| \leq C||(I-P_h)Qx(t)||, \quad t \in J_o, \tag{4.6}$$

where $C$ is independent of $h$ and $t$.

Proof: By applying Theorem 4.1, we obtain the existence of a $C^r$-function $\eta_h: J_o \to Y$ such that $H_h(t,\eta_h(t)) = 0$ for $t \in J_o$ and

$$||\eta_h(t)-\eta(t)|| \leq C_o||H_h(t,\eta(t))||, \quad t \in J_o, \tag{4.7}$$

where $C_o$ depends only on $\gamma_o$ and $\gamma_1$ and hence is independent of $h$ and $t$. With $x_h(t) = x_o + tz_o + A\eta_h(t)$, we have $\hat{F}_h(x_h(t)) = H_h(t,\eta_h(t)) = 0$ for $t \in J_o$. Proposition 3.1 (i) shows that $x_h(t) \in E_h$ and $F_h(x_h(t)) = y_{oh}$; that is, $x_h: J_o \to E_h$ is a solution segment of the local approximate problem (3.5). Finally, the estimate (4.6) follows from (4.7) and

$$||x(t)-x_h(t)|| = ||A(\eta(t)-\eta_h(t))|| \leq ||A||\ C_o||H_h(t,\eta(t))||$$

$$= C_o||A||\ ||(I-P_h)Qx(t)||, \quad t \in J_o.$$

It is worth noting that Lemma 4.4 implies a uniformity in this estimate. In other words, we have the following result:

Corollary 4.6: Under the conditions of Theorem 4.5, we have

$$x(t) = \lim_{h\to 0} x_h(t) \text{ uniformly for } t \in J_o. \tag{4.8}$$

## 5. Mildly Nonlinear Problems

The remaining three sections of this paper are devoted to examples of our general theory. We begin by showing that for mildly nonlinear operators our assumptions, and especially condition (C), are naturally satisfied. This covers the case considered in [5].

As noted at the beginning of Section 3, in many applications we find that $X = Y \times \mathbb{R}^1$. Let $Q \in L(X,Y)$ be the natural projection

$$Q(u,\lambda) = u, \quad x = (u,\lambda) \in X. \tag{5.1}$$

Let $\hat{X}$ be another Banach-space, $K \in L(\hat{X},Y)$ a compact mapping, $G\colon X \to \hat{X}$ an operator of class $C^r$, $r \geq 1$, and define

$$F\colon X \to Y, \quad F(x) = Qx + KG(x), \quad x \in X. \tag{5.2}$$

With $x = (u,\lambda)$, our problem (2.2) then takes the familiar form

$$u + KG(u,\lambda) = y_0. \tag{5.3}$$

Many elliptic boundary-value problems can be written in the form (5.3). For example, consider the problem

$$\begin{aligned} Lu + g(u,\lambda) = -\sum_{i,j=1}^{n} \frac{\partial}{\partial \xi_i}\left(a_{ij}(\xi)\,\frac{\partial u}{\partial \xi_j}\right) + g(u,\lambda) = 0 \quad \text{in} \quad \Omega, \\ u = 0 \quad \text{on} \quad \partial\Omega, \end{aligned} \tag{5.4}$$

where $\Omega$ is a suitable bounded domain in $\mathbb{R}^n$, the coefficients $a_{ij}$ are sufficiently smooth, and the linear part $L$ is strongly elliptic. A weak formulation of (5.4) is

$$\int_{\Omega} \left[ \sum_{i,j=1}^{n} a_{ij}(\xi) \frac{\partial u}{\partial \xi_j} \frac{\partial w}{\partial \xi_i} + g(u,\lambda)w \right] d\xi = 0, \quad w \in H_0^1(\Omega). \tag{5.5}$$

Under appropriate regularity conditions, any weak solution $u \in H_0^1(\Omega)$ of (5.5) turns out to be a strong solution of (5.4). In order to write (5.5) in the form (5.3), suppose that $g$ defines a $C^r$-mapping, $r \geq 1$, $G: H_0^1(\Omega) \times \mathbb{R}^1 \to H^{-1}(\Omega)$. For the strongly elliptic bilinear form

$$a(u,w) = \int_{\Omega} \sum_{i,j=1}^{n} a_{ij}(\xi) \frac{\partial u}{\partial \xi_j} \frac{\partial w}{\partial \xi_i} d\xi, \quad u,w \in H_0^1(\Omega), \tag{5.6}$$

let $K: H^{-1}(\Omega) \to H_0^1(\Omega)$ be the linear operator defined by

$$a(Ku,w) = (u,w)_0, \quad u \in H^{-1}(\Omega), \quad w \in H_0^1(\Omega), \tag{5.7}$$

where $(u,w)_0$ is the inner product on $L^2(\Omega)$. Then $K: L^2(\Omega) \to H_0^1(\Omega)$ is compact and the weak formulation (5.5) is equivalent to (5.3) for $y_0 = 0$.

As in Section 3, we introduce a family $\{P_h\}$ of projections $P_h \in L(Y)$ of finite rank such that (3.3) holds and define the subspaces $Y_h = P_h Y$, $X_h = Y_h \times \mathbb{R}^1$, and $Z_h = Y_h \times \{0\}$, $h > 0$.

With $\hat{F}_h$ as defined in (3.7), the following lemma is the key to the applicability of our results to (5.3):

Lemma 5.1: For the problem (5.3) and for any fixed $x_0 = (u_0,\lambda_0) \in E$, we have $\lim_{h \to 0} D\hat{F}_h(x_0) = DF(x_0)$ in the uniform operator topology on $L(X,Y)$.

Proof: Evidently, the definitions of the operators $F$ and $\hat{F}_h$ imply that $DF(x_0) - D\hat{F}_h(x_0) = (I-P_h)KDG(x_0)$, and the compactness of $K$ ensures that $(I-P_h)KDG(x_0) \to 0$ in the uniform operator topology on $L(X,Y)$.

Next, let $\{V,A,z_0\}$ be a local parametrization of $M(y_0)$ at a point $x_0 = (u_0,\lambda_0) \in M(y_0)$. Then we can verify the following key fact:

<u>Lemma 5.2</u>: For problem (5.3), condition (C) is satisfied.

<u>Proof</u>: Once again we proceed as in the proof of Proposition 3.5. Clearly, $DF(x_0)A$ is an isomorphism of $Y$, and hence the result is a direct consequence of the fact that, by Lemma 5.1, $\lim_{h\to 0} D\hat{F}_h(x_0)A = DF(x_0)A$ in the uniform operator topology on $L(Y)$.

With this, we have shown that, for mildly nonlinear operators, our basic assumptions indeed are satisfied and hence that Theorems 3.7 and 4.5 apply. We note, in particular, the form of the estimate (4.6) in this case. If, as in Theorem 4.5, $x(t) = (u(t),\lambda(t))$, $t \in J_0$, and $x_h(t) = (u_h(t),\lambda_h(t))$, $t \in J_0$, denote the points on the solution segments of the original and approximate problems, respectively, then (4.6) becomes

$$||u(t)-u_h(t)|| + |\lambda(t)-\lambda_h(t)| \leq C||(I-P_h)u(t)||, \quad t \in J_0. \tag{5.8}$$

All points on regular solution curves may be classified as being either nonsingular points or limit points. It is interesting to see what types of local parametrizations correspond to these two types of points as considered in the first two parts of [5].

In defining nonsingular points and limit points, there is no need to restrict ourselves to mildly nonlinear mappings.

<u>Definition 5.3</u>: Let (A) be given and suppose that a splitting (3.1) is available

together with the operator $Q$ of (3.2). Then -- with a mild abuse of notation -- a point $x_0 = (u_0, \lambda_0) \in E$, $u_0 \in Z$, $\lambda_0 \in T$, is <u>nonsingular</u> if $D_uF(x_0)$ is an isomorphism of $Z$ onto $Y$. The point is a <u>limit point</u> (with respect to the direction of $T$) if $D_uF(x_0)$ has a one-dimensional null space in $Z$ and

$$D_\lambda F(x_0) \notin \text{rge } D_uF(x_0). \tag{5.9}$$

The regularity of such points is the content of the following result:

<u>Proposition 5.4</u>: Let the information of Definition 5.3 be given. If $x_0 \in E$ is either a nonsingular point or a limit point, where $DF(x_0) \in L(X,Y)$ is a Fredholm operator of index 1, then $x_0 \in R(F)$.

<u>Proof</u>: If $x_0$ is a nonsingular point, then, since $D_uF(x_0)$ maps onto $Y$, we have $\text{rge } DF(x_0) = Y$, and, since $DF(x_0)$ is a Fredholm mapping of index 1, it follows that $\dim \ker DF(x_0) = 1$ and hence that $x_0 \in R(F)$. If $x_0$ is a limit point, then the formula $DF(x_0)(w,\mu) = D_uF(x_0)w + \mu D_\lambda F(x_0)$ for $(w,\mu) \in X$, together with (5.9), implies that $\ker DF(x_0) = \ker D_uF(x_0) \times \{0\}$, and thus $\dim \ker DF(x_0) = \dim \ker D_uF(x_0) = 1$. Since $DF(x_0)$ is a Fredholm operator of index 1, it follows that $\text{rge } DF(x_0) = Y$ and again that $x_0 \in R(F)$.

We return again to the mildly nonlinear case (5.2). Clearly, since $Q$ is a Fredholm mapping of index 1, so is $DF(x) = Q + KDG(x)$ for any $x \in X$. Thus, Proposition 5.4 applies. If $x_0$ is a nonsingular point, then, because of $(Y \times \{0\}) \cap \ker DF(x_0) = \{0\}$, we may choose as a local parametrization the subspace $V = Y \times \{0\}$, the isomorphism $A: Y \to V$ defined by $Aw = (w,0)$, $w \in Y$, and $z_0 = (0,1)$. This parametrization amounts to the selection of $\lambda$ as the parameter and corresponds in essence to the choice in the first part of [5]. The estimate

(5.8) now reduces to

$$||u(\lambda)-u_h(\lambda)|| \leq C||(I-P_h)u(\lambda)||, \quad \lambda \in J_0. \tag{5.10}$$

If $x_0 \in X$ is a limit point, then with $u_0 \in Y$ such that $D_uF(x_0)u_0 = 0$, $||u_0|| = 1$, and $Y_0 = \text{span}\{u_0\}$, $Y_1 = \text{rge}\, D_uF(x_0)$, it follows tnat $Y = Y_1 \oplus Y_0$, $D_uF(x_0)$ is an isomorphism of $Y_1$, and $\ker DF(x_0) = Y_0 \times \{0\}$. Since $(Y_1 \times \mathbb{R}^1) \cap \ker DF(x_0) = \{0\}$, we may use the local parametrization specified by $V = Y_1 \times \mathbb{R}^1$, the isomorphism $A: Y \to V$ defined by $Aw = (w_1, t_0)$ for $w = w_1 + t_0u_0 \in Y$, $w_1 \in Y_1$, and $z_0 = (u_0, 0)$. This corresponds exactly to the approach in the second part of [5].

## 6. More General Nonlinear Problems

The techniques and results presented here provide a unification which is certainly of interest in itself, but their real value should derive from the fact that the results are applicable to a wider class of problems than the mildly nonlinear problems to which other known approaches appear to be restricted. The full extent of this applicability is still under study, but our results certainly cover problems which to our knowledge could not be handled by other methods. A model problem of this type is discussed in this section.

The flexibility in our formulation lies in the freedom of choice of the operator $Q$ of (3.2). In the case of the mildly nonlinear operators discussed in the previous section, $Q$ may be chosen as the natural projection (5.1). For more general operators, $Q$ has to reflect more specifically some of the properties of the problem. The flexibility in the choice of $Q$ may even be important in the case of mildly nonlinear boundary-value problems such as (5.4).

Although (5.4) can be transformed into a mildly nonlinear problem, the techniques to be discussed below also allow for a direct use of the strong form (5.4).

As our model problem, we consider here

$$\frac{d}{d\xi} \arctan\left(\frac{du}{d\xi}\right) + g(u,\lambda) = 0, \quad 0 < \xi < 1,$$
$$u(0) = u(1) = 0, \tag{6.1}$$

where $g: \mathbb{R}^2 \to \mathbb{R}^1$ is a sufficiently smooth function. A problem of this type was used in [3] as an example-case for numerical experimentation with problems concerning the deformation of nonlinear rods.

An attempt to analyze (6.1) by means of a weak formulation as in the previous section will run into difficulties because the resulting operator $F$ does not satisfy the required continuity and Lipschitz conditions. This is due to the nonlinearity arising from the arctan function. In order to work with (6.1) directly, let $Z = C_0^2[0,1]$ and $Y = C^0[0,1]$, and consider the operator

$$F: X = Z \times \mathbb{R}^1 \to Y, \quad F(x) = d[\arctan(du)] + g(u,\lambda), \quad x = (u,\lambda) \in X, \tag{6.2}$$

where, for abbreviation, we use $du = \frac{d}{d\xi} u$. The derivative of $F$ is given by

$$DF(x)(w,\mu) = d\left[\frac{1}{1 + (du)^2} dw\right] + D_u g(u,\lambda)w + \mu D_\lambda g(u,\lambda), \quad x = (u,\lambda) \in X,$$
$$(w,\mu) \in X. \tag{6.3}$$

It is a straightforward (but somewhat tedious) matter to show that $F$ is in-

finitely differentiable and DF is Lipschitz-continuous on bounded subsets in the sense of (4.5).

For $G: X \to Y$, $G(x) = g(u,\lambda)$, $x = (u,\lambda) \in X$, it follows from Ascoli's Theorem that, for fixed $x_0 = (u_0,\lambda_0) \in X$, $DG(x_0) \in L(X,Y)$ is compact. In view of this and (6.3), we now define

$$Q \in L(X,Y), \quad Q(w,\mu) = d\left[\frac{1}{1 + (du_0)^2}\, dw\right], \quad (w,\mu) \in X. \tag{6.4}$$

It is easy to prove the following two properties of $Q$:

$$\begin{aligned} &\text{(i)} \quad \ker Q = \{0\} \times \mathbb{R}^1, \\ &\text{(ii)} \quad Q|Z \times \{0\} \text{ is an isomorphism from } Z \times \{0\} \text{ onto } Y. \end{aligned} \tag{6.5}$$

In terms of $Q$ and $DG$ the derivative (6.3) of $F$ at $x_0$ has the form

$$DF(x_0) = Q + DG(x_0). \tag{6.6}$$

Now proceed as before and let $\{P_h\}$ be a family of projections $P_h \in L(Y)$ of finite rank and define the subspaces (3.4). With $\hat{F}_h$ as in (3.7), we have $DF(x_0) - D\hat{F}_h(x_0) = (I-P_h)DG(x_0)$, and as in Section 5 we can show that $\lim_{h\to 0} D\hat{F}_h(x_0) = DF(x_0)$ in the uniform operator topology on $L(X,Y)$. It follows in turn that, for any local parametrization $\{V,A,z_0\}$ of $M(0)$ at a point $x_0 = (u_0,\lambda_0) \in M(0)$, the condition (C) holds.

Clearly now, for the mapping $F$ of (6.2) and with $Q$ defined by (6.4), our results apply to the problem (6.1). If $x(t) = (u(t),\lambda(t))$, $t \in J_0$, and $x_h(t) = (u_h(t),\lambda_h(t))$, $t \in J_0$, once again denote the points on the solution segments of the original problem and the approximate problem, respectively, then the estimate (4.6) takes here the form

$$\|u(t)-u_h(t)\|_{C^2} + |\lambda(t)-\lambda_h(t)| \leq C\left\|(I-P_h)\, d\left[\frac{1}{1+(du_o)^2}\, du(t)\right]\right\|_{C^0}, \quad t \in J_o. \tag{6.7}$$

The mapping $F$ of (6.2) serves only as an example of the class of nonlinear operators to which the approach in this section may be applied. In principle, let $F$ be any mapping of the form

$$F: X = Z \times \mathbb{R}^1 \to Y, \quad F(x) = N(x) + G(x), \quad x \in X. \tag{6.8}$$

If $x_o \in X$ is a point for which $Q = DN(x_o)$ has the properties (6.5) and $DG(x_o)$ is compact, then clearly the results apply.

We note in passing that the general discussion of nonsingular points and limit points at the end of Section 5 and the corresponding parametrizations carries over to mappings of the form (6.8).

## 7. The Reduced-Basis Technique

Our final example concerns the so-called reduced-basis technique which has been receiving increased attention in the engineering literature (e.g., see the survey [11] where other references are also given). The technique arises in the context of a standard continuation procedure for nonlinear problems and is usually viewed as a method for effecting considerable reductions in the size of the systems of equations obtained from finite-element approximations. Alternately, as discussed in [6], the technique corresponds to the construction of a set of finite-element basis functions which closely reflects the properties of the solution segment that is to be approximated. In this form, the reduced-basis technique turns out to fit very naturally into the general setting of this paper.

As noted earlier, condition (C) establishes a certain compatibility between the discretization and local parametrization. In the previous two sections, we saw that its validity can be ensured for certain classes of operators without restricting the choice of the discretization and parametrization. Conversely, it is possible to enforce condition (C) without placing any further conditions on the mapping $F$ by selecting certain types of discretizations and parametrizations inherent to the problem. The reduced-basis technique represents an example of this latter approach.

As usual, suppose that the information (A) is given and, for ease of notation, assume that the mapping $F$ is defined on all of $X$. For a fixed $y_o \in F(R(F))$ and $x_o \in M(y_o)$ we consider the mapping

$$F_o: X \to Y, \quad F_o(x) = F(x_o+x), \quad x \in X, \tag{7.1}$$

and note that $0 \in M_o(y_o) = F_o^{(-1)}(y_o) \cap R(F_o)$. Now our procedure is applied to $F_o$ rather than $F$.

For a splitting (3.1) of $X$, we choose $X = Z \oplus T$, where $T = \ker DF(x_o)$ and $Z$ is any complementary subspace. Then $Q = DF(x_o) \in L(X,Y)$ satisfies $\ker Q = T$, and $Q|Z$ maps $Z$ isomorphically onto $Y$. As usual, let a family $\{P_h\}$ of finite-rank projections of $Y$ be given for which (3.3) holds and define the subspaces (3.4). For a local parametrization $\{V,A,z_o\}$, we now choose $V = Z$, $A = A_V = (Q|Z)^{-1}$, and any nonzero vector $z_o \in T$. Then we have

$$\begin{aligned} D\hat{F}_{oh}(0)A &= (I-P_h)QA + P_h DF_o(0)A \\ &= (I-P_h)DF(x_o)A + P_h DF(x_o)A \\ &= DF(x_o)A_V = I, \end{aligned}$$

where $I$ is the identity on $Y$. In other words, condition (C) holds trivially, regardless of the nature of $F$.

Now Theorem 4.5 applied to $F_o$ ensures the existence of an approximate solution segment $x_{oh}(t) = 0 + tz_o + A\eta_h(t) \in X_h$, $t \in J_o$, satisfying $P_hF_o(x_{oh}(t)) = y_{oh}$, $t \in J_o$. Therefore, $x_h^R(t) = x_o + x_{oh}(t)$ lies in $x_o + X_h$ and satisfies

$$F_hF(x_h^R(t)) = y_{oh}, \quad t \in J_o. \tag{7.2}$$

From the estimate (4.6), we see that $x_{oh}(0) = 0$ and thus $x_h^R(0) = x_o$. By differentiation of (7.2) with respect to $t$ at $t = 0$, it follows that

$$0 = P_hDF(x_o)(z_o + A\eta_h'(0)) = P_h\eta_h'(0). \tag{7.3}$$

Since $z_o$ and $tz_o + A\eta_h(t)$ are in $Y_h$, the same holds for $\eta_h(t)$, $t \in J_o$, whence (7.3) implies that $\eta_h'(0) = 0$ or $\frac{d}{dt}\, x_h^R(0) = z_o$.

Hence, the solution segment $x_h^R(t)$, $t \in J_o$, passes through $x_o$ and has the same tangent direction as the original curve. By appropriate choice of the space $Z$ and the projections $P_h$, it is easy to ensure that higher derivatives of the original curve and the approximate curve also agree at $x_o$. (Note, however, that the validity of condition (C) requires only that $\ker DF(x_o) \subset X_h$ and not any additional matching of higher derivatives at $x_o$.) This is precisely the approach of the reduced-basis technique; that is, the method uses the first few directions of the moving frame of the original solution curve $M_o(y_o)$ at the origin as the basis vectors for the approximating subspaces $X_h$. For some details and an error assessment, we refer to [6]. As the mentioned survey [11] already indicates, the method is certainly beginning to prove itself very

effective for the solution of various geometrically nonlinear structural problems.

## References


[1] J. C. Alexander and S. S. Antman, Global behavior of solutions of nonlinear equations depending on infinite-dimensional parameters, Technical Report, Department of Mathematics, University of Maryland, 1981.

[2] K. E. Atkinson, The numerical solution of a bifurcation problem, SIAM J. Numer. Anal., 14, 1977, 584-599.

[3] I. Babuška and W. C. Rheinboldt, Computational error estimates and adaptive processes for some nonlinear structural problems, Univ. of Pittsburgh, Inst. for Comp. Math. and Applications, Tech. Report ICMA-82-28, 1981, to appear in Comp. Meth. in Appl. Mech. and Eng.

[4] Yu. G. Borisovich, V. G. Zvyagin and Yu. I. Sapronov, Nonlinear Fredholm maps and the Leray-Schauder theory, Russian Math. Surveys, 32:4, 1977, 1-54.

[5] F. Brezzi, J. Rappaz and P.-A. Raviart, Finite-dimensional approximation of nonlinear problems.
Part I: Branches of nonsingular solutions, Numer. Math., 36, 1980, 1-25.
Part II: Limit points, Numer. Math., 37, 1981, 1-28.
Part III: Simple bifurcation points, Numer. Math., 38, 1981, 1-30.

[6] J. P. Fink and W. C. Rheinboldt, On the error behavior of the reduced-basis technique for nonlinear finite element approximations, Univ. of Pittsburgh, Inst. for Comp. Math. and Applications, Tech. Report ICMA-82-37, April 1982, to appear in Zeit. f. Angew. Math. und Mech.

[7] V. Girault and P.-A. Raviart, An analysis of upwind schemes for the Navier-Stokes equations, SIAM J. Numer. Anal., 19, 1982, 312-333.

[8] H. B. Keller, Numerical solution of bifurcation and nonlinear eigenvalue problems, in "Applications of Bifurcation Theory", ed. by P. Rabinowitz, Academic Press, New York, 1977.

[9] R. Lozi, A computing method for bifurcation boughs of nonlinear eigenvalue problems, Bull. Amer. Math. Soc., 81, 1975, 1127-1129.

[10] J. W. Milnor, Topology from the Differentiable Viewpoint, The University Press of Virginia, Charlottesville, Virginia, 1965.

[11] A. K. Noor, Recent advances in reduction methods for nonlinear problems, Computers and Structures, 13, 1981, 31-44.

[12] G. W. Reddien, Computation of bifurcation branches using projection methods, SIAM J. Math. Anal. and Appl., 81, 1981, 204-218.

[13] M. Schechter, Principles of Functional Analysis, Academic Press, New York, 1971.

[14] G. Vainikko, Approximative methods for nonlinear equations (two approaches to the convergence problem), Nonlin. Anal., Theory, Methods, and Appl., 2, 1978, 647-687.

[15] R. Weiss, Bifurcation in difference approximations to two-point boundary-value problems, Math. Comp., 29, 1975, 746-760.

DATE
FILMED

11-82

DTIC